90-27

Axial Anomalies *

L.O'Raifeartaigh

Dublin Institute for Advanced Studies,
10 Burlington Road, Dublin 4, Ireland.

Abstract

Anomalies are chiral symmetries of classical Lagrangians that are broken at the quantum level. They occur only to first order in $\hbar$ (for topological reasons) but are of central importance for particle physics. The aim of these lectures is to explain the origin of the anomalies and to describe their theoretical and experimental consequences.

1.Introduction.

It is well-known that most of the symmetries of classical physics are also symmetries of quantum physics, and this is particularly true of the space-time symmetries such as Lorentz symmetry. However, there is at least one classical space-time symmetry, namely chiral symmetry, which is not a symmetry of the quantum theory,and the breaking of this symmetry at the quantum level,and its consequences, are the subjects that I wish to consider in these lectures. In terms of the Lagrangian, what happens is that if the effective Lagrangian is written as

$$L_{eff} = L_c + L_{\triangle q}, \tag{1}$$

where L_c is the classical Lagrangian and $L_{\triangle q}$ is its quantum correction, then $L_{\triangle q}$ does not inherit the chiral symmetry of L_c. More precisely, if L_c contains fermion fields $\psi(x)$ then $L_{\triangle q}$ does not inherit the symmetry of L_c with respect to the transformations

$$\psi(x) \rightarrow e^{i\gamma_5 \alpha(x)}\psi(x), \tag{2}$$

of these fields. One may consider both the rigid (or global) chiral transformations, $\alpha(x) = \textit{constant}$, and the local chiral transformations, $\alpha(x) \rightarrow 0$ as $x \rightarrow \infty$. The experimental cosequences of the breakdown of the rigid chiral symmetry are (i) the fact that the number of lepton and quark families must be the same (ii) the associated evidence for the existence of just three colours for the quarks (iii) the resolution of the so-called $U(1)$-problem (which is posed by the observed breakdown of the chiral part of the central $U(1) \times U(1)$ subgroup of the flavour internal symmetry group $U(n) \times U(n)$), and

Proc. Paul Scherrer Institute Spring School, Zuoz, Switzerland 1990

(iv) the light that the resolution of the $U(1)$ problem throws on the problems of the mass and three-pion decay of the η particle. Experimental consequences of the breakdown of local chiral symmetry are (i) the prediction of the correct rate for the decay $\pi \to 2\gamma$ (ii) the associated evidence for the existence of just three colours for the quarks and (iii) the prediction of a decay-rate for baryons, which, although miniscule at present temperatures, might have been appreciable in the early universe. Finally it might be mentioned that analogous chiral anomalies occur in gravitational, supergravitational, Kaluza-Klein and String theories, and the condition that their coefficients vanish (in order to make the theories consistent with renormalization) put restrictions on these theories. For example the vanishing of the coefficient of the conformal anomaly in string theory leads to the well-known condition that the dimension of the configuration space be 26 or 10.

2.Description of Anomaly.

As the anomaly actually occurs only to first order[**1**] in $\hbar$ it is possible to discuss it purely within the context of the WKB approximation[**2**]. Let us begin by recalling the (time-dependent) WKB approximation in ordinary Quantum Mechanics with Schrodinger equation

$$\frac{\hbar}{i}\frac{\partial}{\partial t}\psi(x) = [\frac{\hbar^2}{2m}\triangle + V]\psi(x). \tag{3}$$

If we write

$$\psi(x) = \rho^{\frac{1}{2}} exp(\frac{i}{\hbar}S), \tag{4}$$

and take the real and imaginary parts we obtain

$$\frac{\partial}{\partial t}S + \frac{1}{2m}(\nabla S)^2 + V = O(\hbar^2), \tag{5}$$

and

$$\hbar[\frac{\partial}{\partial t}\rho + \frac{1}{m}\nabla(\rho\nabla S)] = 0. \tag{6}$$

These equations have two remarkable features: (a) In eq.(6) Planck's constant appears only as an overall factor so that the equation itself is classical. It is actually the continuity equation

$$\frac{\partial}{\partial t}\rho = divj, \quad \text{where} \quad j = \frac{1}{2m}\bar{\psi}\overleftrightarrow{\nabla}\psi, \tag{7}$$

which expresses the conservation of probability, in disguise. (b) In the WKB approximation equation (5) is classical, in fact it is just the Hamilton-Jacobi equation for the corresponding classical system. Furthermore,if one uses the solution S_{HJ} in (6) one can solve locally for ρ, and the solution with the correct initial condition to describe a particle located at $x = x(o)$ at time t=0 is simply

$$\rho(x) = det[\frac{\partial^2 S_{HJ}}{\partial x_a(t)\partial x_b(o)}]^{-1}. \tag{8}$$

Thus finally the WKB approximation for the solution $\psi(x)$ to the time-dependent Schrodinger equation is

$$\psi(x)_{WKB} = det[\frac{\partial^2 S_{HJ}}{\partial x_a(t)\partial x_b(o)}]^{-\frac{1}{2}} \exp(\frac{i}{\hbar}S_{HJ}). \tag{9}$$

Thus in the WKB approximation $\psi(x)$ is expressible completely in terms of the classical action $S_{HJ} = \int L_c dt$. For example, for a free ($V = 0$) particle in n dimensions $\psi(x)$ is just $\frac{1}{(mt)^{n/2}}\exp(i\frac{(x-a)^2}{2\hbar mt})$.

What has all this got to do with anomalies? First write (9) as

$$\exp\{\frac{i}{\hbar}[S_{HJ} - \frac{\hbar}{2i}\ln\det\frac{\partial^2 S_{HJ}}{\partial x_a(t)\partial x_b(o)}]\}. \tag{10}$$

This can be interpreted as

$$S_{WKB} = S_{HJ} - \frac{\hbar}{2i} ln det[\frac{\partial^2 S_{HJ}}{\partial x_a(t)\partial x_b(o)}], \tag{11}$$

the second term giving the first-order QM corrections to the classical term S_{HJ}. If S_{HJ} is invariant with respect to any space-time symmetries such as rotations then, since the x's are space-time vectors and occur only in the determinant, the correction term is also rotationally invariant. In other words, in the QM case the quantum correction inherits the space-time symmetry of the classical term S_{HJ}. The point now is that in Quantum Field Theory (QFT) this is not always the case. In particular in QFT involving fermions and chiral space-time symmetries it is generally not the case. In these cases the analogue of the Hamilton -Jacobi term S_{HJ} is the Dirac or Yang-Mills action

$$S_{HJ} = \int dt L_c = \int d^4x L_c(x) = \int d^4x \bar{\psi}_c(x)[i\not{D} + m]\psi_c(x), \tag{12}$$

where subscript c denotes classical and $\not{D} = \gamma^\mu D_\mu$, where D_μ is the covariant derivative i.e. $D_\mu = \partial_\mu + ieA_\mu$ for electromagnetism and $D_\mu = \partial_\mu + \sigma_a A^a_\mu$, where the σ's are the group generators, for nonabelian Yang-Mills theory.What we need now is the first-order quantum correction to (12).This is computed [**3**] from the (exact) functional integral formula

$$\exp[S_{eff}] = \exp[S_{HJ}]\int d\psi d\bar{\psi} e^{S(\bar{\psi},\psi,A)}, \tag{13}$$

which, by the rules of fermionic functional integration [4], is just

$$S_{eff} = S_{HJ} + \frac{\hbar}{i} lndet[\frac{\delta^2 S(\bar{\psi}, \psi, A))}{\delta\bar{\psi}(x)\delta\psi(x)}]. \tag{14}$$

Note the resemblance between this result and the QM one. (The change in sign and factor 2 comes from the fact that the integration is fermionic and that there are two fermion fields).For the Dirac or Yang-Mills cases, where S is linear in $\bar{\psi}$ and ψ, it is clear that the quantum correction is just

$$\frac{\hbar}{i}[lndet(i\not{D} + m)], \tag{15}$$

(where it is understood that the *lndet* has to be regularized).The statement of the anomaly now is that with respect to the chiral transformations

$$\psi(x) \to e^{i\gamma_5\alpha(x)})\psi(x), \tag{16}$$

and the induced transformations for $\not{D}$ and m, the quantum correction (15) is not invariant. This is, perhaps, not too surprising because, on account of the γ_o involved in the definition of $\bar{\psi}$, the latter transforms in the *same* way as ψ (not in the opposite way) and thus the functional derivative $\delta\bar{\psi}\delta\psi$ in (14) is not chirally invariant. Note that, on account of the same factor γ_o in the definition of $\bar{\psi}$ the induced transformations of $\not{D}$ and m just mentioned are

$$\not{D} \to \not{D} + [\not{D}, \gamma_5\alpha(x)]_+ \quad \text{and} \quad m \to m + 2m\gamma_5\alpha(x), \tag{17}$$

for infinitesimal α. Note also that the statement that the functional derivative with respect to $\bar{\psi}$ and ψ is not chirally invariant is equivalent to Fujikawa's well-known observation [**3**] that the measure $d\bar{\psi}d\psi$ in the functional integral is not chirally invariant, since for fermionic variables integration and differentiation are the same! The quantitative form of the statement that the quantum correction (15) is not chirally invariant is that

$$\frac{\delta}{\delta\alpha}[\frac{\delta^2 S_{HJ}}{\delta\bar{\psi}\delta\psi}] = \frac{\delta}{\delta\alpha}[lndet(i\not{D} + m)] = \frac{e^2}{16\pi^2}E.B, \tag{18}$$

where

$$E.B \equiv F^{\mu\nu}F^*_{\mu\nu} \equiv \epsilon_{\mu\nu\lambda\sigma}F^{\mu\nu}F^{\lambda\sigma}.$$

Here **E** and **B** are the electric and magnetic fields (and their Yang-Mills equivalents) and the quantity $E.B$ is a pseudoscalar, as it should be, since the parameter α is a pseudoscalar.

3.Derivation of Formula (18).

For those interested in how the result (18) actually emerges let me just sketch the idea. Those not interested can safely ignore this section. First, one regularizes $lndet(i\not{D}+m) = Trln(i\not{D}+m)$ by the ζ-function method, i.e. by writing it as

$$\frac{1}{2}Tr(-\not{D}^2+m^2) = -[\frac{d}{ds}\zeta(s)]_{s=o}, \quad \text{where} \quad \zeta(s) = \frac{1}{2}Tr(-D^2+m^2)^{-s}. \tag{19}$$

Next, using (17) and the fact that, since, for sufficiently large s the quantity $(-D^2+m^2)^{-s}$ is trace-class one can cyclically permute operators ,one finds that within the trace the chiral variation of $(-\not{D}^2+m^2)^2$ reduces to $4\alpha(x)\gamma^5$ times itself. Thus the chiral variation of $\zeta(s)$ is

$$\delta\zeta(s) = 2(-s)Tr[(-\not{D}^2+m^2)^{-s}\alpha(x)\gamma_5],. \tag{20}$$

which can also be expressed in the integral form

$$\delta\zeta(s) = \frac{-2s}{\Gamma(s)}\int dt t^{s-1}e^{-m^2t}Tr[\alpha(x)\gamma_5 e^{(-t\not{D}^2)}]. \tag{21}$$

It is not difficult to see that, provided that the trace is a regular function of t, its contribution to $\delta\zeta'(s)$ at $s=0$ comes only from its value at $t=0$. So all one has to do is to verify the regularity of the trace as a function of t and compute it at $t=0$. This is easily done by noting that $\not{D}^2 = D^2+\sigma_{\mu\nu}F^{\mu\nu}$, where the sigmas are the Lorentz generators, and that, to leading order in t, (i) $e^{\not{D}^2} = e^{tD^2}e^{t\sigma_{\mu\nu}F^{\mu\nu}}$ (ii) $tr[(\gamma_5)\exp(\sigma.F)] = E.Bt^2$ where tr denotes the Dirac trace, and (iii) $\text{Trexp}(-tD^2) = \text{Trexp}(-t\triangle) = O(t^{-2})$, where $\triangle$ is the free Laplacian and its proportionality to t^{-2} follows from the earlier expression for the free wave-function in n dimensions with n set equal to 4. Putting these four observations together one sees that the trace is indeed regular, and that its value at $t=0$, and hence the required value of the anomaly, is just $E.B$ up to a numerical constant.

4.Violation of Current Conservation.

The most convenient way to connect the breakdown of chiral symmetry to physical quantities is to express it in terms of a breakdown of current conservation. It is well-known that any (rigid) continuous symmetry of a (Lagrangian) theory leads to the conservation of a local current (Noether's theorem) $J_\mu(x)$ say, and a more precise statement of this result (due to Noether herself) is that

$$\partial_\mu J_\mu(x) = \frac{\delta L(x)}{\delta\alpha(x)}, \tag{22}$$

where $\alpha(x)$ is the (infinitesimal) symmetry group parameter. For rigid chiral symmetry, the current $J_\mu(x)$ computed according to the usual Noether rules is $\bar{\psi}(x)\gamma_\mu\gamma_5\psi(x)$ and the variation of the Lagrangian contains two terms, namely a "classical"term $2m\bar{\psi}(x)\gamma_5\psi(x)$ coming from the variation of m in (17) and the anomalous term (18). [Note that from rigid chiral symmetry i.e. $\alpha(x)$=constant, the variation of $\not{D}$ in (17) is zero, so there is no contribution from this term]. Thus from the chiral symmetry (or, more precisely, lack of symmetry) of the Lagrangian we have

$$\partial_\mu\bar{\psi}(x)\gamma_\mu\gamma_5\psi(x) = 2m\bar{\psi}(x)\gamma_5\psi(x) + \frac{e^2}{8\pi^2}E.B. \tag{23}$$

This equation shows how the anomaly affects current conservation, and since it is the currents that connect directly to physics, how it affects physical quantities. For example, since at low energies the neutral pion field $\pi^o(x)$ can be approximated by the axial current by means of the Golberger-Treiman [**5**] relation

$$\pi^o(x) = (m_\pi^2 f_\pi)^{-1}\partial_\mu J_\mu^5(x), \tag{24}$$

where $J_\mu^5(x)$ is the current on the left-hand-side of (23), m_π is the pion mass, f_π is the weak coupling constant determined from β-decay, one sees how, in principle, (24) can determine the decay rate for $\pi^o \to 2\gamma$. (More details will be given in section 7).

5.General Prefactor.

The above description was for an axial field coupled to a single fermion field but it is clear that it can be generalized to an axial field coupling to any number of fermion fields by replacing the prefactor e^2 by $tr(\sigma_5 Q^2)$ where σ_5 and Q denote the axial and electric charges of the fermions respectively, and the trace is with respect to these internal indices. Note that in this case the chiral gauge transformation is

$$\psi(x) \rightarrow e^{(i\sigma_5 \alpha(x))}\psi(x). \tag{25}$$

More generally, if the EM field is replaced by any Yang-Mills field (including the axial field itself) the anomaly generalizes to

$$tr(\sigma^5 \sigma_a \sigma_b) E^a . B^b, \tag{26}$$

where the $\sigma_{a,b}$ are the generators of the Yang-Mills group. The generalization to the case when the axial field is also non-abelian yields a more complicated expression [**6**] than $E.B$ for the dynamical part of the anomaly but yields the same prefactor, except that σ^5 is replaced by the generators σ^5_a of the non-abelian axial gauge group. The general prefactor

$$tr(\sigma^5_c \sigma_a \sigma_b) \tag{27}$$

is important because the renormalizibility of the theory requires that the anomaly, and hence the prefactor, must vanish when this trace is taken with respect to *all* the fermion fields. We shall see later that for the standard model this is not the case for the leptonic or quark sectors alone but is true for the combined sectors, provided that the numbers of leptonic and quark families are equal.

6. The Triangle Graph.

Historically the anomaly was first discovered [7] in the course of computing the triangle graph of Fig.1, and it is still often associated with that graph. Hence it might be useful to establish the relationship between the anomaly as defined above and that defined by the graph. This is done as follows: First we neglect the mass m as we know in advance that it does not appear in the final expression for the anomaly. Then we return to the original expression for the quantum correction to the Dirac or Yang-Mills Lagrangian i.e. $Trln(i\not{D})$. Using

$$\delta \not{D} = [\not{D}, \gamma_5 \alpha(x)]_+ = 2\gamma_5\gamma_\mu\partial_\mu\alpha(x), \tag{28}$$

we see at once that

$$\delta Tr \ln \not{D} = 2Tr[\gamma_5\gamma_\mu\partial_\mu\alpha(x)(\not{D})^{-1}]. \tag{29}$$

Now expanding $(\not{D})^{-1}$ in the form

$$(\not{D})^{-1} = (\not{\partial})^{-1} + (\not{\partial})^{-1}e\not{A}(\not{\partial})^{-1} + (\not{\partial})^{-1}e\not{A}(\not{\partial})^{-1}e\not{A}(\not{\partial})^{-1} + ... \tag{30}$$

and inserting the expansion in (29) we see that the Dirac trace kills the two leading terms and so

$$\delta Tr \ln \not{D} = 2Tr[(\not{\partial})^{-1}e\not{A}(\not{\partial})^{-1}e\not{A}(\not{\partial})^{-1}\gamma_5\gamma_\mu\partial_\mu\alpha(x) + ...]. \tag{31}$$

It is possible to see that the higher order terms represented by the dots do not contribute, either from the fact that the higher order terms in t in the previous derivation do not contribute, or from the fact that the corresponding Feynman graphs are finite. Hence the only contribution to the anomaly comes from the term displayed in (31) and it is evident that this term is just the one represented by the triangle graph of Fig.1. Note that since γ_5 commutes with $(\not{\partial})\alpha(x)(\not{D})^{-1}$ the anomaly is just the difference of the anomalies due to the left-handed and right-handed fermions i.e.the two kinds of fermions do not mix in the loop of Fig.1. Note also that the general prefactor (28) can be read off from the triangle graph since the three $\sigma's$ are just the usual internal symmetry insertions at the three vertices.

7. Π°-Decay.

The most important application of the local anomaly is to the decay of the π^o meson into two photons. This process depends on the fact that at low energies the π^o field can be approximated by the divergence of the axial vector current. More precisely, at low energies one has the Goldberger-Treiman relation (24). It follow that the amplitude for the decay $\pi^o \to 2\gamma$ can be written as

$$< \pi^o|2\gamma >= (m_\pi f_\pi)^{-1} < 0|(\partial_\mu j^5_\mu)^o|2\gamma > . \tag{32}$$

However, from the theory of PCAC (partially conserved axial vector currents) it can be shown [**5**][**8**] that the first (mass) term on the right-hand side of (23) does not contribute to this process. Indeed, before the discovery of the anomaly, this was one of the few PCAC results in disagreement with experiment and was regarded as a major puzzle. The puzzle was resolved by the anomaly, in the form of the second term on the right-hand side of (23), which shows that if the first (mass) term is neglected in (23) the amplitude for the decay is given by

$$< \pi^o|2\gamma >= (16\pi^2 m_\pi^2 f_\pi)^{-1} tr(\sigma_3^5 Q^2) < o|F^{\mu\nu}F^*_{\mu\nu}|2\gamma >, \tag{33}$$

where σ_3^5 is twice the neutral generator of chiral isospin, and Q the electric charge, of the participating fermions. If one denotes by k_μ, k'_ν and e_λ, e'_σ the momenta and polarization vectors of the emerging photons then the amplitude (33) takes the form

$$< \pi^o|2\gamma >= (l6\pi^2 m_\pi^2 f_\pi)^{-1} tr(\sigma_3^5 Q^2)\epsilon_{\mu\nu\lambda\sigma} p^\mu p'^\nu e^\lambda e'^\sigma. \tag{34}$$

Since the decay process takes place in the strangeness zero sector and is a low energy effect the fermions that effectively participate are the proton and neutron and hence the trace factor in (34) is effectively

$$tr(\sigma_3^5 Q^2) = Q_p^2 - Q_n^2 = Q_p^2 = 1. \tag{35}$$

Inserting this result in (34) one obtains a decay rate for the π^o which is in excellent agreement with experiment. An interesting feature of this result is that it was obtained [**9**] as far back as 1949 by Steinberger, who proposed (33) using other considerations, and carried out the subsequent computations. An even more remarkable feature of the result, however, is the one described in the next section.

8.The Number of Quark Colours.

We have just seen that the π^o decay rate comes out correctly if we use the anomaly and regard the nucleons as the participating fermions. On the other hand, according to the quark model, the nucleons are composed of the up- and down-quarks, so we should obtain the same result if we use these quarks instead of the nucleons. In that case the trace factor becomes

$$tr(\sigma_3^5 Q^2) = N_c(Q_u^2 - Q_d^2) = N_c[(2/3)^2 - (1/3)^2] = N_c/3, \qquad (36)$$

where N_c denotes the number of colours, and we see that this will give the same result (unity) as the nucleons if, and only if, N_c is equal to three. Thus, through the anomaly, the π^o decay rate gives us a direct measurement of the number of colours. This is an important result because the original statistics argument for the existence of colours requires only that there must exist at least three colours,and the π^o measurement is one of the few pieces of evidence that there are just three colours. In fact there are really only three pieces of direct experimental evidence for three colours, namely the π^o decay just discussed, the vanishing of the anomaly for the present electroweak generations to be discussed in the next section, and the cross-section ratio $R = \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)}$ for electron positron collisions, which depends [**10**] strongly and monotonically on the number of colours.

9.Renormalizability and Generations.

In this section we indicate why renormalizability requires the vanishing of the anomaly when the trace in the prefactor is taken over *all* fermion fields and how this in turn requires that the number of lepton and quark families be the same. We first note (from the triangle graph for example) that the anomaly is proportional to the momenta of the external gauge fields. In the case that these gauge fields are external fields for a complete Feynman graph this is not an immediate problem (though ultimately it could cause problems with the unitarity bound). But if the triangle is in the interior of some higher order Feynman graph then the momenta of the gauge fields are *internal* momenta and so increase the divergences of the usual integrals over momentum. Since the divergence of these integrals is such that they are renormalizable, but not superrenormalizable or finite, even in the absence of these extra momentum factors, it is not difficult to imagine that they become non-renormalizable in their presence, and this is exactly what happens [**11**]. So if we wish to have a renormalizable theory we have to make sure that, when the anomaly appears in the interior of a Feynman graph, it vanishes. On the other hand, when the anomaly does appear as the internal part of some Feynman graph then the fermion fields participating in it consist of *all possible* such fields, not a selected subset as in the case of the nucleons or up- and down-quarks as in π^o decay. Hence for renormalizability we need only require that the prefactor trace taken over *all* fermion fields vanish,

$$tr_{\Sigma}(\sigma_c^5 \sigma_a \sigma_b), \tag{37}$$

where the subscript Σ means that the trace is the complete one i.e. is with respect to all fermion fields. If we recall that the left- and right-handed fermions do not mix, but contribute seperately to the anomaly, this condition may be written a little more explicitly as

$$tr_{\Sigma}(\sigma_c^5 \sigma_a \sigma_b)_L = tr_{\Sigma}(\sigma_c^5 \sigma_a \sigma_b)_R. \tag{38}$$

In evaluating (38) we first note that if the assignment of the fields to the representations of the gauge group is vectorial i.e. is such that the left- and right-handed components of a field are assigned to the same representation of the gauge group, then the field will give no contribution to the anomaly. This is what actually happens for the strong interactions, where the assignment of the quarks to the (fundamental) representation of the colour gauge group is indeed vectorial. We should also note that particles and anti-particles have opposite handedness, and that if a particle is assigned to a representation of a gauge group its anti-particle is assigned to the complex conjugate representation. It follows that for particles belonging to representations of the gauge group that are equivalent to their complex conjugates the contribution of the particle to the left-hand-side of (38) is matched by the contributions of its anti-particle to the right-hand side, so the contribution of such a pair to the anomaly is zero. Hence the only fields that we really have to worry about are those that are assigned to representations that are not equivalent to their complex conjugates. Such (unsafe) representations exist for only a limited number of groups, namely the (unsafe) groups $U(1), SU(n), n \geq 3, SO(4n+2)$ and $E(6)$, and even for these groups not all of the representations are unsafe [**12**]. (In fact, for $SO(4n+2)$ and $E(6)$ only the representations congruent to the spinor and 27-dimensional representations are unsafe). However, unsafe representations occur in the assignments of the standard model, and that is where they introduce restrictions that relate the number of quarks to the number of leptons. We recall that the gauge group for the standard model is $SU(2)\times U(1)$ (more precisely $U(2) = SU(2)\times U(1)/Z_2$). Since $SU(2)$ is a safe group there is no contribution to the anomaly coming from the $SU(2)$ part alone so the problem comes from the pure $U(1)$ and mixed $U(1)$ -$SU(2)$ contributions.From the triangle graph one sees that if Y and σ denote the generators of $U(1)$ and $SU(2)$ respectively the prefactors for the anomalies coming from these two sources are

$$tr(Y^3) \quad \text{and} \quad tr(Y\sigma^a\sigma^a), \tag{39}$$

respectively.(There is no anomaly with Y^2 in the trace because the σ and Y traces are independent and the trace of a single $SU(2)$ generator σ is zero). We now recall that in the standard model [**13**] the weak hypercharge assignments of the left-handed quarks and leptons are

$$Y = (-1,-1), 0, 2, (1/3, 1/3), -4/3, 2/3, \tag{40}$$

where the brackets denote isospin doublets and the particles are ordered in the conventional way, e.g. for the first generation the ordering is

$$(\nu^e, e), \bar{\nu}^e(?), \bar{e}, (u,d), \bar{u}, \bar{d}. \tag{41}$$

We also recall that since Y generates a $U(1)$ group the generator of a complex conjugate represention is (-Y). Hence for Y the contributions of the left-handed particles and right-handed anti-particles to the anomaly will not cancel but add up. Hence if the anomaly is to vanish the contribution of the left-handed particles alone must vanish. It is clear from (40) (and from the fact that $tr(\sigma^a\sigma^a)$ is 3/4 for doublets and 0 for singlets) that

$$\begin{aligned} trY^3 &= (6) + N_c(-54/9) = 2(3-N_c) \\ tr(Y\sigma^a\sigma^a) &= (-2) + N_c(2/3) = (-2/3)(3-N_c), \end{aligned} \tag{42}$$

where the quark contributions can be identified as those proportional to N_c (which, as before, denotes the number of colours). From this equation we see that neither anomaly is zero for the quarks and leptons alone, but that both anomalies vanish if (and only if) both the quarks and leptons of each seperate generation are included,and if the number of colours is three. Thus the vanishing of the anomaly, as required by renormalizability, requires that each set of standard model leptons be accompanied by a set of standard model quarks. Thus, via the anomaly, renormalizability requires that the number of quark and lepton families must be the same. This means, for example, that the recent experiments at CERN which established that there are only three families of (low-mass)leptons thereby automatically established that there are also only three families of (low-mass) quarks. This correlation of quarks and leptons which is obtained via the anomaly is most remarkable as it is the first such correlation between hadronic and leptonic physics that has ever been derived, and it cannot be obtained in any other way. In conclusion, it might also be noted that for the gravitational anomaly, in which the σs in (39) are replaced by the universal gravitational constant which may be normalized to unity, the prefactor reduces to trY, and that this vanishes for the quarks and leptons seperately. Thus even the lepton and quark sectors seperately of the standard model have no gravitational anomaly.

10.Resolution of the $U(1)$ Problem.

Another use of the anomaly is to resolve a long-standing problem in flavour-symmetry theory called the $U(1)$ problem.This problem arises as follows:In the zero quark mass limit the flavour symmetry of the strong interactions is $G = U_L(n) \times U_R(n)$, where L and R refer to left- and right-handedness, and n=2,3,4...for the various flavour groups i.e. n=2 for isospin, n=3 for Gell-Mann's $SU(3)$, n=4 for Charm, and so on. Since the $U(n)$ groups are not simple but have a centre $U(1)$ the group G can also be written as $G = U(1)_L \times U(1)_R \times SU(n)_L \times SU(n)_R$ (modulo some discrete global correlations that are not important here). Since chirality doubling is not observed in nature (there are no 0^+, $\frac{1}{2}^-$, $\frac{3}{2}^-$ octets to match the 0^-, $\frac{1}{2}^+$, $\frac{3}{2}^+$ octets in the Rosenfeld tables, for example) one concludes that the axial part of the group G is spontaneously broken, leaving only the the vector (diagonal) part $U(1)_v \times SU(n)_v$, where $U(1)_v$ is the symmetry that conserves baryon number and $SU(n)_v$ is the original vector version of the isospin,Gell-Mann $SU(3)$, Charm $SU(4)$...groups. However, as is well-known, the breakdown of a continuous symmetry entails the existence of either (massless) Goldstone fields or (massive) gauge fields to cancel them by the Higgs mechanism. For the breakdown of the chiral part of $SU(n)_L \times SU(n)_R$ it is assumed that the observed pseudo-scalar mesons are the required Goldstone fields (the pions for n=2, the pion,kaon and η_8 fields for n=3, and so on, their observed non-zero masses, which are small on the relevant scale, being acquired afterwards by some other process associated with the generation of the quark masses). But for the $U(1)_L \times U(1)_R$ part of the group no such Goldstone field or compensating Higgs gauge field has been seen experimentally. It might be thought that the η_o particle could be such a Goldstone field but estimates from current algebra show [**14**] that the mass later acquired by any such Goldstone field should not exceed $\sqrt{3}m_\pi$, whereas the mass of the η_o exceeds m_π by factor of about four. So the η_o is ruled out, and one is left with the puzzle: How does $U(1)_L \times U(1)_R$ break down to $U(1)_v$ without producing a Goldstone scalar or a Higgs vector? This is the $U(1)$ problem. The resolution [**15**] of this problem by the anomaly comes at two levels. Assuming that the breakdown of $U(1)$ chiral symmetry is due to the anomaly according to (23), then at the first level, one notes that there is a sense in which the chiral symmetry is not broken at all. This is because the anomalous term in (23) can also be expressed as a divergence,

$$E.B = F^*_{\mu\nu}F^{\mu\nu} = \partial_\mu J^A_\mu, \quad \text{where} \quad J^A_\mu = \epsilon_{\mu\nu\lambda\sigma}[A_\nu F_{\lambda\sigma} + \frac{1}{3}A_\nu A_\lambda A_\sigma], \tag{43}$$

and hence if this term is shifted to the left-hand side of the equation one obtains a conservation law for the modified axial current

$$\partial_\mu(J^5_\mu - cn_f J^A_\mu) = 0, \tag{44}$$

where c is a numerical constant and n_f is the number of fermion fields. Since the modified axial $U(1)$ symmetry is not broken there is then no need for Goldstone scalars or Higgs vectors.

At the second level, however, one sees that, while one has solved the Goldstone-Higgs aspect of the problem in this manner, one has created another problem, because the question now is: why is the conserved axial $U(1)$ symmetry not observed in nature in the same way as the conserved vector $U(1)$ symmetry e.g. as the conservation of left- and right-handed fermions seperately? The answer to this problem is that the piece which has been added to the traditional fermionic axial current in (43) is (manifestly) not gauge-invariant and hence the symmetry it generates is not physical. This explanation may seem a little facile but it can be put on a somewhat more quantitative basis by considering the conserved axial charge G_5 and its action on the vacuum. Let $G_5^A(t)$ denote the part of G_5 coming from the gauge field alone. It is clear from (43) that G_5^A is not gauge- invariant, and is also not seperately conserved since

$$G_5^A(\infty) - G_5^A(-\infty) = \int d^4x F^*_{\mu\nu} F^{\mu\nu} = I, \tag{45}$$

where I is the (integer-valued) instanton number. Hence the eigenvalues of $G_5^A(t)$ are integer-spaced and without loss of generality can be taken to be integers n. The vacuum can then be decomposed into n-states i.e. into states $|n>$ such that

$$G_5^A(t)|n>= n|n> \quad \text{and} \quad I = n(\infty) - n(-\infty) \tag{46}$$

This implies, of course that the vacuum is degenerate. The states $|n>$ are not gauge-invariant or time-independent because $G_5^A(t)$ itself does not have these properties. However, because the n's change only by integers it is easy to see that the linear combinations

$$|\theta>= \Sigma_n e^{in\theta}|n>, \tag{47}$$

of these states are both gauge-invariant and time-independent for each value of the arbitrary parameter θ. Since the states which are linear comnbinations of these states with different values of θ are, like the $|n>$ states, neither gauge-invariant nor time independent, this means that there must be a superselection rule which divides the vacuum, and hence the whole theory, into gauge-invariant sectors [16] parametrized by θ. The transformations generated by the conserved axial charge violate this superselection rule since they change the θ-states into one another thus are not physical transformations.

In sum therefore, the accepted resolution of the $U(1)$ problem is that the axial current is actually conserved, but that the axial charge is gauge-variant and therefore does not generate physically observable transformations. As this explanation may seem to be a little unusual it might be useful to consider a symmetry of classical physics that could, perhaps, be regarded as analogous. This is the dual symmetry

$$\mathbf{E} \to \mathbf{E}\cos\alpha + \mathbf{B}\sin\alpha, \quad \mathbf{B} \to \mathbf{B}\cos\alpha - \mathbf{E}\sin\alpha, \tag{48}$$

of the free Maxwell equations, which is continuous and conserved, but has no direct physical meaning.

It is only fair to mention, of course, that this resolution of the $U(1)$ problem raises other questions, namely: what determines the value of θ occurring in nature? and, since this value is extremely small ($\theta \neq 0$ implies CP-nonconservation because the instanton number I in (45) is a pseudoscalar, and is therefor ruled out by the neutron dipole moment experiments), what makes it so small? But, in contrast to the case of the original $U(1)$ problem, there is no inherent contradiction involved here, and these are simply open questions.

Finally it should be mentioned that although the $U(1)$ symmetry is not directly observable it may be indirectly observable in the form of the η-mass and the decay of the η to three pions. The idea is that the η may acquire its mass by an analogue of the Higgs mechanism for gauge fields or fermions and the $\eta \to 3\pi$ decay may proceed by an analogue of the $\pi_o \to 2\gamma$ decay. However, the arguments for these processes (for example the argument for the η-mass given in the next section) are not as rigorous or as convincing as those for the processes already discussed, so perhaps it would be best to refer here to the literature, for example to the books [**14**], [**15**] and [**17**] listed below.

11.Fermion-Instanton Interactions.

It is now well-known that the non-abelian Yang-Mills field equations admit solutions $F_{\mu\nu}$ for which the classical Euclidean action $S_c(F) = \int d^4x F^{\mu\nu} F^*_{\mu\nu}$ is finite but not zero (the so-called instanton solutions). The Minkowski-space interpretation of these solutions is that they represent a tunneling between the degenerate eigen-vacua $|n>$ of the operator $G_5^A(t)$ defined in the previous section. Indeed, from (45) one sees that in the presence of an instanton $G_5^A(\infty) - G_5^A(-\infty)$ is not zero so that the eigenvalues n must change with time. From the usual Quantum Mechanical WKB formula the amplitude for the tunnelling is then found to be $\exp(-S_c)$.

The relevance of all this to the anomaly is that through the conservation equation (23) the instantons couple to fermions and lead to an effective fermion-fermion interaction of a kind that would not otherwise be there. It is this induced fermion-fermion interaction that is thought to lead to a mass and a 3π decay of the right order of magnitude for the η mentioned above. Here, however, we wish to discuss a more dramatic, if less observable, consequence of instanton-fermion coupling, namely baryon-decay. The point is that the instantons do not distinguish between different baryons and hence the leptons and quarks can couple via the instantons. In an attempt to make the argument more quantitative 't Hooft [**18**] adopted the following approach. Let $j_{r,s}$, where r and s are flavour indices, denote a set of external chiral (left-handed, say) currents coupling to the fermions through the usual kind of term

$$\bar{\psi}_r(1-\gamma_5)\psi_s j_{r,s}, \tag{49}$$

in the Lagrangian, and consider the first quantum corrections to the classical action $S_c(F)$ of a background instanton. Since $j_{r,s}$ behaves like a mass parameter except for its space-time dependence one sees from (14) that the action up to first order in $\hbar$ is

$$S_{WKB} = S_c + \frac{\hbar}{i} Trln[i\not{D} + (1+\gamma_5)j_{r,s}], \tag{50}$$

where $\not{D}$ is the covariant derivative for the gauge-potential $\not{A}$ of the background instanton and the change in sign in front of γ_5 comes from the γ_o in the definition of $\bar{\psi}$. The most important contribution to the Trace in (50) comes from the zero modes of $\not{D}$ and thus is proportional to the currents, and a more precise calculation shows that it takes the form

$$\Pi_{r,s} \frac{(1+\gamma_5)j_{r,s}}{(x-x_o)^6}, \tag{51}$$

where x_o is the 'centre-of-mass' of the instanton. 't Hooft then asked the question: what fermion-fermion interaction would produce (51)? By using fermionic functional integration theory it is not difficult to see that the answer is

$$\Pi_{r,s}\bar{\psi}_r(1+\gamma_5)\psi_s, \tag{52}$$

and from this 't Hooft concluded that (52) is the effective fermion-fermion interaction induced by the instantons. [Note that the induced fermionic interaction has the right quantum numbers to generate an η mass, and that for two flavours it even takes the form of a mass-term, except for its space-time dependence]. Here, however, we are interested in baryon-decay. It is clear that (52) makes no distinction between different baryons and hence allows baryons to couple to leptons without any selection rules apart from chirality. Indeed the only thing that prevents an immediate decay of all baryons into leptons is the exponential tunnelling factor $\exp(-S_c)$, which is extremely small. For example, for a single instanton $S_c = (16\pi^2 g^2)^{-1}$ so it is of the order of 10^{-150} and thus 10^{120} smaller than the corresponding factor for the grand-unification decay rate. This factor, however, is the one computed at zero temperature, and temperature-dependent computations show that once the temperature rises above the tunnelling-temperature i.e. the temperature needed for direct transitions from one n-vacuum to another, the situation changes dramatically. This will be discussed in the lecture by Andreas Wipf, who will show that the temperature in question can be estimated by means of sphalarons, and that it is only of the order of 200 Gev. Thus, in principle, instanton-induced baryon-decays could have taken place in the early universe at any time prior to the time of the electroweak breakdown.

References.

[1] W. Bardeen, Phys. Rev. **184** (1973) 1848, L. Alvarez-Gaume Comm. Math. Phys. **90** (1983) 161.

[2] D. Amit, Field Theory, Renormalization Group and Critical Phenomena, McGraw- Hill, New York (1978).

[3] K. Fujikawa, Phys. Rev. **D21** (1980) 2848.

[4] P. Ramond Field Theory, A Modern Primer, Benjamin-Cummings, New York (1981).

[5] S. Adler, R. Dashen, Current Algebra and Applications to Particle Physics, Benjamin, New York (1968).

[6] S. Treiman, R. Jackiw,B. Zumino,E. Witten, Current Algebra and Anomalies, Princeton University Press (1985).

[7] S. Adler, Phys. Rev. **177** (1969) 2426,
J. Bell, R. Jackiw, Nuovo Cim. **60A** (1969) 47.

[8] B. Renner, Current Algebra and Applications, Pergamon, New York (1968).

[9] E. Leader, E. Predazzi, Gauge Theory and the New Physics, Cambridge University Press (1982).

[10] I. Aitcheson, A. Hey, Gauge Theory in Particle Physics, Hilger, New York (1982).

[11] D. Gross, R. Jackiw, Phys. Rev. **D6** (1972) 477.

[12] L. O'Raifeartaigh, Group Structure of Gauge Theories, Cambridge University Press (1988).

[13] P. Becher, M. Bohm, H. Joos, Eichtheorien der Starken und Electroschwachen Wechselwirkungen, Teubner Stuttgart (1981). q